A Guide to System Data

MassDOT/MBTA Office of Performance Management and Innovation (OPMI)

Included in This Guide

Key Words, pg. 1-4
How We Collect System Data, pg. 5-8
Vehicle Location Data Use Cases, pg. 9
Reference Table Use Cases, pg. 10

Key Words

System Data

OPMI defines system data as files that describe the MBTA including where, when, or under what circumstances the system is operating. Many of the data in this category report service statistics, facts, and references. Several files are supplemental and have expected uses such as mapping or joining. System data files may include data on headways, travel times, GTFS, service area, reference topics, and more.

Vehicle Location Data

We have vehicle tracking on almost all our vehicles, with the exception of the Mattapan Line trolleys and the commuter ferries. With this basic "vehicle *v* in place *p* at time *t*" data we can derive headways and travel times for performance measurement and make predictions to put on signs and in our real-time data feed. Read more about how each mode is calculated here.

GTFS

(General Transit Feed Specification)

The General Transit Feed Specification (GTFS) is a data specification that allows public transit agencies to publish their transit data in a format that can be consumed by a wide variety of software applications. Today, the GTFS data format is used by thousands of public transport providers.

Service Area

The core service area refers to the 64 cities and towns in Massachusetts for which MBTA bus or rapid transit service is provided. The extended service area refers to the original 64 plus the remaining 111 cities or towns that also have MBTA commuter rail service. This data is based off of the 2010 census. Rhode Island cities and towns with MBTA service are not included in the "Service Area."

Key Words

Event
(Rail)

There are four types of event: ‘ARR’ for arrival, and ‘DEP’ for departure, ‘PRA’ for predicted arrival, and ‘PRD’ for predicted departure. In the last two cases, the system uses the last available prediction for the given event as a reasonable proxy.

Travel Time
(Rail)

The actual time it takes for a train to travel between the selected origin stop and the destination stop, in seconds. Calculated using rail events.

Headway
(Rail)

Time between the departure time of the current trip and the departure time of the previous trip for trips of the same “destination”, for rapid transit vehicles. Red line trunk stops will have two headways for the same southbound train: one dependent on the destination and one independent of the destination. Calculated using rail events.

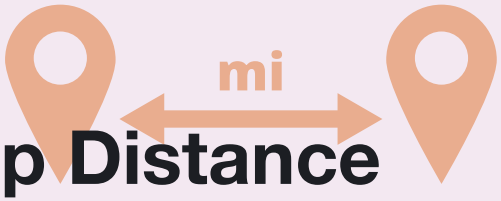
Headway
(Bus)

The actual time between the trip and the previous trip at the stop, in seconds. Endpoints are evaluated by whether the trip runtime is within 120% of the scheduled run time.

Key Words

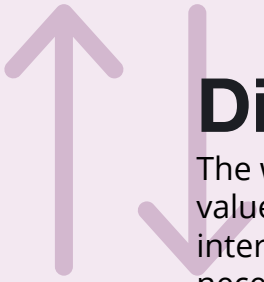
Alerts

Notifications of altered service or schedules. Alerts may be planned or result from unforeseen circumstances. Current and planned alerts are available here. Historical alerts are currently available using the MBTA API. An example of the call and its inputs can be viewed here.



Stop Distance

The distance between stations, in miles. This is not the length of track segments between stations.



Direction

The way in which a vehicle is traveling. In our data, direction is indicated by a binary value, a direction plus “bound”, or the destination terminal. These are not always interchangeable, and the binary value assigned to the direction of a train does not necessarily correlate to inbound/outbound. For example, a Red Line Braintree train at Davis on the inbound track has three directions: id (0), description (Southbound), and terminal station (Braintree). On physical train cars and buses, the direction is indicated by the headsign showing the destination of the vehicle.

Stop Sequence

The order in which stops and stations are placed along a route. Differs by direction and route variants.

Key Words

Peak Hour

According to the MBTA Service Delivery Policy, there are two distinct intervals for “rush hour.” Morning rush hour, or AM_PEAK, is from 7:00 am to 8:59 am, and evening rush hour, or PM_PEAK, is from 4:00 pm to 6:29 pm.

Time Period

According to the MBTA Service Delivery Policy, there are nine time periods for service on weekdays, shown in the table to the right. Saturdays and Sundays are measured separately for most standards.

Time Period	Times
Very Early Morning	3am-5:59am
Early AM	6am-6:59am
AM Peak	7am-8:59am
Midday Base	9am-1:29pm
Midday School	1:30pm-3:59pm
PM Peak	4pm-6:29pm
Evening	6:30pm-9:59pm
Late Evening	10pm-11:59pm
Night	12am-2:59am

Service Day

An MBTA service date is defined as 3:00 am – 2:59 am (2:59 am the next day).

Season

Most of our ridership files contain averages from our fall rating. Ratings are the four times a year that the MBTA makes major schedule changes, but are not necessarily the same as the calendar seasons. Fall is generally September through December.

How We Collect System Data

There are several ways in which we gather data about the service provided by the MBTA. Most of these data sets are spatial in some way, whether they map service availability or track actual movements of vehicles. Therefore, we depend on different sources to create files for the Open Data Portal. Below are two of the methods we employ.

Using Vehicle Location Data

Knowing MBTA vehicles' locations is critical to ***manage our service***, ***measure our performance***, and provide passengers with ***predictions of upcoming vehicle arrivals***. How we get that information varies among the different modes, and each tracking technology has different advantages and limitations that affect how we handle it.

Mattapan Line

In recent years, GPS tracking has been enabled on the Mattapan line, providing the system with the vehicle number in transit. As with above ground Green line stations, the Mattapan trolleys are tracked using pings every specified interval to geolocate and infer station arrival and departure times.

How We Collect System Data

Using Vehicle Location Data

Heavy Rail

For heavy rail, train locations come from the signal system, which also enforces safe distances between trains. The subway tracks are divided into “blocks” that are a few hundred feet long (at least the length of a full train).

When a train is in a block it triggers a track circuit that indicates to the signal system that that block is occupied. This indication of block occupancy is sent back to the control center and also recorded in a database. The database knows which blocks are stations, which tells us when a train enters and leaves a station by recording when it entered and left that block.

Bus

Vehicle locations on buses are part of a broader vehicle technology suite called TransitMaster that the MBTA uses to manage bus service. TransitMaster uses a combination of GPS and the vehicles’ odometer to determine the bus’ location.

When a bus reaches certain stops that we call time points, this data is recorded in a performance management server that compares it to the schedule and determines if the bus was on time. The real-time information and performance reporting programs get bus locations at least once every minute and get additional data when buses pass time points.

How We Collect System Data

Using Vehicle Location Data

Green Line

Since the Green Line has elements of light rail, streetcar and subway modes, we need to combine different data sources and methods to know where its vehicles are. Further details on Green Line tracking are available here.

1 Track Circuits

A track circuit detects the presence (or absence) of a train on a particular section of track. Unlike an AVI, which is located at a specific, known point in space, a track circuit is a linear segment that could be hundreds of yards (even miles) long, although on the Green Line they are typically much shorter than that. Only some sections of the Green Line have track circuits.

2 Automatic Vehicle Identifiers (AVI)

The AVI system includes sensors on the tracks called “key points” that read information from Green Line cars that pass by. When AVI equipment detects the presence of a train car, it communicates with the car via RFID and gets three pieces of information:

- The ID number of the car (these are painted on the inside and outside of all Green Line cars)
- The ID number of the route that the car is on
- Whether or not the car is coupled to another car

3 GPS

GPS fills in the gaps of location knowledge for aboveground stations. A GPS device gives us the location and ID number of the train car that it is on, but no route data.

How We Collect System Data

Using GTFS

We publish system data from GTFS and GTFS archives. The current feed is available to the public here as a zip. The MBTA GTFS feed is updated at least four times a year but is subject to change at any time. These four times refer to the four MBTA ratings, or seasons when service changes. However, changes made to the feed outside of these scheduled times does happen. We use some of the required GTFS tables to determine our reference tables. There are optional and experimental tables included in the GTFS zip as well. Further explanation and documentation are available on the MBTA Github.

GTFS Table Name	Includes
Agency	agency_ids, names, time zones, languages
Calendar	service_ids, day of the week binary identifiers, start dates, end dates
Routes	route_ids, names, descriptions, vehicle types
Stops	stop_ids, names, descriptions, latitudes, longitudes, parent stations
Stop Times	trip_ids, arrival times, departure times, stop_ids, stop sequences
Trips	route_ids, service_ids, trip_ids, headsigns, directions, route patterns

We also have a systemwide GTFS map file available here.

The map service contains layers from the stops, shapes, and facilities files from the MBTA GTFS.

Vehicle Location Data Use Cases

Use Case	File Name	File Type	Years	Update Frequency	Aggregation
Find the average and spread of headways for heavy rail and light rail routes and stops Ex. *What is the median headway at South Station for a Braintree-bound Red Line train in Fall 2019?*	Average Rail Headways Fall 2019	Hosted Table	2019	Annually	Day type, peak indicator, route, stop
Find the average and spread of travel times for heavy rail and light rail routes and adjacent stops Ex. *What is the average time to travel between South Station and Braintree on a Saturday in Fall 2019?*	Average Rail Travel Times Fall 2019	Hosted Table	2019	Annually	Day type, peak indicator, route, stop segments
Find actual headway calculations for heavy rail and light rail routes and stops. User must calculate averages, sums, etc. Ex. *What is the average headway at South Station for a Braintree-bound Red Line train for January 2020?*	Rapid Transit Headways	Hosted Table	2016-2020	Quarterly	Time, DD, day type, route, stop
Find actual travel times calculations for heavy rail and light rail routes and stop combinations. User must calcualte averages, sums, etc. Ex. *What is the average time to travel between South Station and Braintree in January 2020 during peak vs off-peak times?*	Rapid Transit Travel Times	ZIP File	2016-2020	Quarterly	Time, DD, day type, route, stop
Find timestamps, counts, and more for any rail event. User must calculate averages, sums, etc. Ex. *How many departures took place at South Station for trains headed southbound?*	Rapid Transit Events	ZIP File	2016-2020	Quarterly	Time, DD, day type, route, stop
Find timestamps, counts, and more for any bus event. User must calculate averages, sums, etc. Ex. *How many departures took place at South Station for the outbound 7?*	Bus Arrival Departure Times	ZIP File	2018-2020	Quarterly	Time, DD, day type, route, stop

Reference Table Use Cases

Use Case	File Name	Updated	Columns Provided
Find the descriptions and terminal stations for binary directions ids Ex. *What is the terminal station for a Red Line train at South Station with direction 1?*	MBTA Rapid Transit Directions	Spring 2020	route, direction id, direction description, terminal station
Find the order of stops by joining on route, stop, and direction or terminal station Ex. *What are the stops direction before and after South Station on a Red Line train heading towards Alewife from Braintree?*	MBTA Rating Transit Stop Orders	Spring 2020	route, direction id, origin, destination, stop order, stop
Find the distance between stations Ex. *What is the distance between South Station and Alewife?*	MBTA Rapid Transit Stop Distances	Spring 2020	route, origin, destination, “from station”, “to station”, distance between (mi), cumulative distance (mi)
Find the exact dates which make up the winter, spring, summer, and fall MBTA ratings Ex. *For which months is the MBTA summer rating used?* Ex. *What is the start and stop of each fall rating over the past three years?*	MBTA Ratings & Seasons	Spring 2020	season name, date start, date end
Find the time frame of all 11 MBTA time periods Ex. *What times are considered peak and which are off-peak?*	MBTA Service Delivery Policy Time Periods	Spring 2020	day type name, time period, period start time, period end time